

AUTOMATED ACUITY SCORING WITHIN A COMPUTER BASED MEDICAL RECORD

Kevin Fitzpatrick, Devin Satz, William Hammond
Ruby Grewal, Edward Hammond III, James Collins, Joseph A. Moylan

Departments of Medical Informatics, General and Thoracic
Surgery, Duke University Medical Center
Durham N.C.

*Supported in part by NLM grant G08LM04513

ABSTRACT

This paper describes the initial development of a completely automated acuity scoring system that resides within the TMR bedside computing system at the Duke University Medical Center, Surgical Intensive Unit. The scoring system is based upon the APACHE II acuity scoring system and provides for the recalculation of acuity scoring at 12 hour intervals through the patient's ICU course. When comparing hand calculated versus computer generated acuity scores for 19 patients, discrepancies fell into three broad categories: 1) data available to the application differed from that available to the human scorer. 2) apparent transcription errors 3) data items lost or absent from the paper record. It remains to be determined if computer generated acuity scoring provides for a more accurate representation of the patient's acuity.

investment in intensive care units is among the least efficient uses of medical care resources². These costs and assertions has prompted the effort to establish meaningful and reliable severity measurement algorithms.

The effort to establish an objective, reproducible measure of severity began in 1974 with the introduction of the Therapeutic Intervention Scoring System². Patients are indirectly ranked by applying a weighted score to various medical interventions. In creating the TISS scoring system, the originators subjectively grouped each of 216 patients into one of four classes of severity, based upon the patient's perceived hemodynamic stability. The specific diagnosis and actual physiologic condition of the patient are not detailed. The primary short-coming of TISS is that it is a surrogate measure of acuity not a direct measure. Peter Rossi and Howard Freeman³ have stated that a surrogate measure is appropriate if a direct measurement is impossible to obtain or its attainment is prohibitively expensive. The more closely related a surrogate is to its corresponding measure, the more meaningful the interpretations arising from it. Rossi⁴ went on to advise that to insure consistency, single critical care nurse be used to collect all TISS acuity data. Shifting critical care nurses away from patient care to data collection is a costly requirement especially given the current shortage of critical care personnel.

INTRODUCTION

The ability to systematically group patients using severity as the common denominator is motivated by the desire to improve quality assurance programs, compare outcomes among heterogeneous groups of patients, and examine resource utilization throughout a health care system. A significant factor in the treatment of a disease is the severity of the patient's condition when receiving treatment. This fact makes accurate and meaningful severity assessment a crucial part of therapy assessment. Critical care medicine currently consumes approximately 15% of the annual \$15 billion dollars spent for hospital care in the United States¹. In the face of these huge expenditures, some researchers argue that

In 1987 Dr. Michael Shabot automated a facsimile of the TISS system citing manpower requirements of the manual system as a motivating factor^{5,6}. In developing this system, Shabot demonstrated that automated data acquisition and analysis was possible. This work represents a landmark achievement in critical care data management. In the early 1980's, a more direct

measure of patient acuity was introduced, the APACHE I severity scoring system. This system used 34 physiologically based measures, and chronic health points to rank severity. Knaus et al. reported that of 63 patients predicted to die 13 lived (21%) and of 519 patients predicted to live 52 died (10%). The primary organ system responsible for ICU admission must be designated to interpret the meaning of an APACHE score. This specificity represents a major change from TISS system. The system does not account for diagnosis and comorbidity. In 1984 Le Gall et. al described a simplified acute physiologic(SAP) scoring system⁷. Like APACHE, SAP utilizes a 0 to 4 point scoring system based upon a patient's physiologic derangement. Unlike APACHE SAPs methodology restricts data collection to the first 24 hours after ICU admission and reduced are the number of physiologic measures from APACHE's 34 to only 14. The authors reported its sensitivity⁸ at 0.56 and its specificity⁹ at 0.82. This is comparable to APACHE. The advantage of SAP is that it reduced data collection time.

In 1987 APACHE II was released¹⁰. The major difference between APACHE II and SAP is that SAP utilizes ventilation or CPAP where APACHE II utilizes the alveolar arterial oxygen tension difference. The data collection period for APACHE II was reduced from 32 hours to only the first 24 hours from ICU admission.

Severity scoring systems that restrict data collection to the first 24 hours all share one criticism; they do not reflect the dynamic pathophysiologic changes occurring during an ICU stay¹¹. They do have an ability to stratify patients into groups exhibiting comparable risk.

TIME SENSITIVE SEVERITY SCORING SYSTEM

The Duke Time Sensitive Severity Scoring System (DTS4) is a fully automated system utilizing the APACHE II scoring system as its basis. DTS4 and APACHE differ in four specific ways: 1) DTS4 calculates patient severity for each 12 hour period of the first 5 days of SICU admission 2) Besides delivering the APACHE score, the DTS4 score utilizes the worst Glasgow Coma Score for each 12 hour period not the best score utilized by APACHE 3) The system calculates and displays changes in severity for each 12 hour period 4) The system calculates the probability of a patient requiring active therapy based upon the APACHE score. This

implementation runs under Database Inc's¹² Total Medical Record (TMR) in the Surgical Intensive Care Unit (SICU) at Duke University Medical Center.¹³

TMR FUNCTIONAL OVERVIEW

TMR is a computerized system supporting the full range of information requirements found in patient care. This implementation required a subset of the supported functions. The data required to calculate patient severity is contained in the demographics, laboratory, subjective and physical evaluations, and nursing assessment sections of the patient record. TMR is written in GEMISCH¹⁴. GEMISCH is a high level object oriented language developed to satisfy the information requirements found in medical environments. The system provides multi-user facilities for collecting, retrieving, manipulating, and displaying medically based information¹⁵. The heart of TMR is an interactive data dictionary. The dictionary provides applications with context sensitive data. Each section of the dictionary is individually accessed by applications. The dictionary documents the location and content of all of the system's information. This centralized control of information significantly reduces application coding and maintenance time, eliminates data redundancy, and improves data quality.

The algorithms used to implement DTS4 utilized the most recently published APACHE weights¹⁶. The date and time of SICU admission represent time zero. To analyze a datum the interval between time zero and the time of data generation is determined. A matrix stores all of the data required to calculate severity. Each row corresponds to a unique factor and each column represents a 12 hour period of time. This application can analyze a patient's severity for up to 120 hours from the time of admission. Each 12 hour period is described as a unique group.

VARIABLES AND DATA STORAGE STRUCTURE

Data for calculating the APACHE score is primarily found in two sections of the electronic patient record, subjective and physical findings (SAP) and laboratory studies. A strict APACHE implementation requires that only data created after admission to the SICU be included in the severity score. In our implementation criteria was relaxed for three reasons. First, it is possible for studies ordered prior to SICU admission to be the catalyst for the

transfer. Second, a paper by Dr. Escarce¹⁷ demonstrated that admission source is a significant co-predictor of severity when associated with APACHE. We interpreted this to mean that events occurring prior to ICU admission are important in determining severity. Third, unlike APACHE's snapshot view, this application presents a dynamic view of a patient's progress. This application accepts studies created up to 24 hours before SICU admission.

TRANSFERRING WORST SCORES TO WORST WEIGHT ARRAY

The program expects multiple scores within each time group. The application selects the worst weight within each group using the *worst weights array*. The worst weights array is a 13 x 10 matrix. Rows represent factors and columns represent groups. Each value set in each factor node is individually examined. The score in the factor node is compared to the score in the associated worst weights cell. If the score in the factor node exceeds the score in the worst weights cell it replaces the value in the cell.

ASSESSING THE GLASCOW COMA SCORE (GCS)

Data required to evaluate the GCS is obtained by examining information related to the neurologic exam in TMR's fully automated nursing note. In this application the fundamental elements that comprise the GCS are imbedded in the nursing dictionary, and are charted for each patient at least once every twelve hour shift. Values for each observation are determined by matching the nurses observation to the list of possible observations stored in the dictionary.

OXYGENATION SCORE

To calculate the oxygenation score the 3 x 10 oxygen array, was created. Determining the APACHE value for oxygenation requires evaluating the FIO₂, PO₂ and PCO₂. The program loops through each factor node selecting the worst lab value in each time group. This value is placed into the oxygen array cell representing the parameter and group it came from. As required by APACHE methodology two approaches are used to assess oxygenation points. If FIO₂ is greater than or equal to .50 (mm Hg) then the A-aDO₂ value is used. This

value is calculated with equation 2.

equation 2: Table Value = $[FIO_2 * 713 - PaCO_2 - PaO_2]$ ¹⁸

For values of FIO₂ less than .50 (mm Hg) the PaO₂ value is used. As before severity weights are cataloged in the dictionary. Once the values are calculated the range to which they belong to is determined. Finally the associated weight is placed into the cell of worst weights array representing the correct group.

CREATININE CALCULATIONS

Determining the serum creatinine weight requires an assessment of renal function. In the presence of acute renal failure (ARF) the APACHE weight for serum creatinine is be doubled¹⁹. The program searches the patients' electronic problem list for the presence of ARF.

CALCULATING CHRONIC HEALTH POINTS

Chronic health points are assigned to patients having chronic organ insufficiency or if immunocompromised states are evidenced prior to hospital admission²⁰. The patient receives five points if they are nonoperative or have undergone emergency surgery. The patient receives two points if classified as having an elective procedure.²¹ Calculating chronic health points required three steps. In the first step, the past medical history contained within the patient's electronic problem list is reviewed for the presence of serious pre-existing condition. A chronic health code flag is set when a problem code in the patient record matches a member of the reference list. The application also determines if the patient has a condition or is receiving therapy known to suppress the immune system. The application determines if the patient had surgery by comparing the patient's record to a list of major procedures. If the code for a major procedure is found in the patient record the application determines occurred during this encounter. If the procedure is identified with this encounter, a major procedure flag is set. Without this flag the application skips to the last section where points are assigned.

To calculate the APACHE score each factor's worst weights for the first 24 hours are selected and totaled²². To this value the age points and chronic health points are added. This total is the APACHE II score.

To calculate the time sensitive score, each column is totalled. Age points from the purge array are added to the corresponding group value. Finally, the chronic health points are added. The result is a severity score for each 12 hour period of the first 120 hours since admission to the SICU. These results are presented graphically. (FIG 1)

A second graph is created to communicate the change in the severity score between each 12 hour period. This graph was prepared following the work by Chang²³ which illustrated the relationship between severity score and its rate of change as significant determinants of patient outcome. (FIG 2)

VALIDATION PROCEDURE AND RESULTS

Nineteen patient records were selected. The APACHE score was determined manually and with DTS4. Whenever a disparity in score was found the data was reviewed to determine the cause of the deviation. All program logic flaws were eliminated.

Even after removing all program logic errors a large number of discrepancies remained. These discrepancies seem to fall into three categories: 1) data available to the application differed from that available to the human scorer 2) apparent transcription errors 3) data items lost or absent from the paper record, or misinterpreted by the recorder. We expected the discrepancies to arise from errors in program logic and didn't structure the study to investigate discrepancies arising from manual data abstraction.

Thirteen records had differences in scores attributed to nonuniform availability of raw data. Delays in connecting patients to the vital signs monitor prevented TMR from capturing some pathophysiologic data. Of the 19 records reviewed, at least four had no data for mean arterial pressure, heart rate, or respiratory rate for at least the first 12 hours of their SICU stay. Based upon the scores prepared manually it appears that the paper record did have data for this period of time. In these cases the manually prepared score indicates a physiologic derangement requiring non-zero APACHE weights for one or more of these factors. In the absence of data the computer assigned no points. A second cause of error is the mirror image of the first. In this case the computer seems to have data not present in the patient record. These errors are characterized

by a human assessment indicating a smaller physiologic derangement than evidenced in the computer record. All errors of this type involved the first set of data made available to the computer.

The next discrepancy category involves laboratory data. In the case of four patients, the laboratory data indicates greater deviation from normal than indicated in the manually prepared scores. If the difference exists in the first group of the DTS4 score it may be attributed to our acceptance of laboratory scores from 24 hours prior to SICU admission. Recall that one of the reasons we accept these scores is to permit capture of these larger abnormal values. All four records in this category exhibited a difference in at least one physiological variable in this first time group. Three records also had differences not explained by this data acceptance criteria. In these three cases it appears that the scorer either lacked access to the full set of laboratory data or failed to appropriately score the laboratory results. Five records exhibited differences in the Glasgow coma score (GCS). These errors can greatly affect the overall severity score. Unlike the majority of factors which can contribute only four points GCS contributes up to 15 points. In three cases the manual system calculated Glasgow coma scores of zero while DTS4's scores ranged from 3 to 12. In the fourth case a manual score of 4 was calculated while DTS4 computed a score of 3. In the fifth case a score of 9 was manually determined while the computer determined a score of 0. In this case the data suggests that data from outside the permitted APACHE time window was used in the manual calculation. In all cases the data in the computer record was reviewed. In all cases the score derived by DTS4 was validated based upon the data available to it.

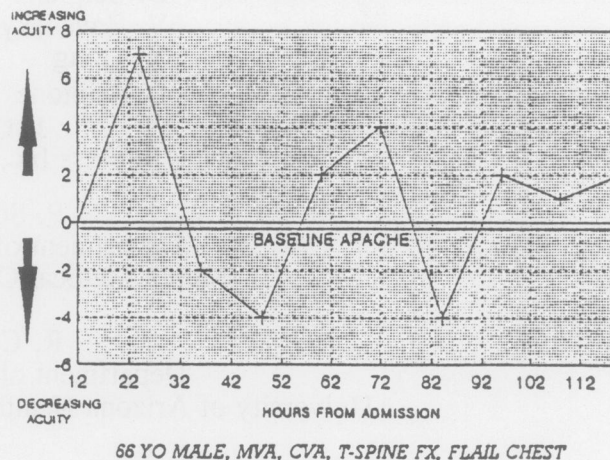
Since its introduction in 1981 the role APACHE has transformed in the literature from a methodology to compare outcomes amongst different institutions to a "prognostic scoring system ... used by physicians as part of their information base for treatment decisions for individual patients"²⁴. All APACHE systems (including the commercially available APACHE III) require clinical personnel manually enter data for analysis. Manual data entry adds the possibility of typographical errors to the set of error types seen in this study.

The affect of delayed vital sign data capture was documented in this study. The average

difference in APACHE score which I've attributed to a delay in data acquisition is 2.0 points. The standard deviation in the difference in points is 0.79 points and the maximum difference for a single factor is 4.0 points. Early in APACHE's development much attention focused on the affect stabilizing a patient prior to admission to intensive care. The Escarce article²⁰ attributes the differences in predicted mortality to the admission source. This study suggests that more to the point than admission source is the affect of time required for initial data collection on point assignment. Before utilizing APACHE methodology as a prognostic device the impact of time delays in data acquisition should be fully understood. This study also indicated that sampling frequency affects severity calculations. Sampling frequency for hemodynamic data in the TMR application is every 5 minutes. In the manual system sampling frequency is every 15 - 60 minutes. It remains to be determined whether transient hemodynamic events that effect the severity scoring is more likely to represent a patient's true condition than one based upon a single hourly sample. It appears reasonable that the optimum sampling frequency is a function of both the patient's severity and the rate of change in their severity.

The automation of acuity indices as part of an automated medical record represents an excellent opportunity to facilitate outcome research, quality assurance and utilization review studies. Automation of these functions will also alleviate the inefficiency associated with assigning highly trained individuals to the task of manual data abstraction. Future research with this system will focus on the optimal sampling frequency of hemodynamic data and the utility of recalculating acuity over time.

DELTA CHANGE IN SEVERITY SCORE AT 12 HOUR INTERVALS



SEVERITY SCORING AT 12 HOUR INTERVALS

